

Данный подход применен в одном из административных органов Республики Беларусь.

Литература

1. *Краснопрошин, В.В.* Интеграция распределенных экспертных знаний: проблемы и решения /В.В. Краснопрошин, Г. Шаках, А.Н. Вальвачев // Информатика. – Минск, 2004. – № 1. – С. 45–53.
2. *Katia, P. Sycara* Multiagent Systems /Katia P. Sycara //AI Magazine. – 1998. – Vol. 10, № 2. – P. 79–93.

ЗАДАЧА О СВЕРТКЕ БЕЛКА НА СПЕЦИАЛЬНЫХ ДВУМЕРНЫХ И ТРЕХМЕРНЫХ РЕШЕТКАХ

Е. А. Левина

Одна из важных задач молекулярной биологии заключается в нахождении третичной структуры белка, исходя из информации об его первичной структуре [1, 2]. Напомним, что упорядоченная последовательность аминокислот, из которых состоит белок, называется первичной структурой белка. Под воздействием физиологических условий каждый белок “свертывается” в уникальную трехмерную структуру, называемую естественной третичной структурой белка. Третичная структура определяет функции, которые белок выполняет в организме. Возникает следующая естественная задача: найти вычислительный метод восстановления третичной структуры белка по известной упорядоченной последовательности аминокислот, ассоциированной с данным белком. В литературе эта задача известна как задача о свертке белка (protein folding problem) [1 – 3].

На практике часто рассматривают упрощенные модели для задачи о свертке белка [4, 5]. Одной из таких моделей является НР-модель, предложенная К. Диллом [5]. В этой модели белок трактуется как последовательность аминокислот, рассматриваемых с точностью до гидрофобности, а пространство свертки описывается с помощью геометрического графа, ассоциированного с некоторой двумерной или трехмерной решеткой. Свертке белка соответствует инъективное отображение аминокислот в вершины графа, при котором соседние в последовательности аминокислоты соответствуют смежным вершинам графа. Третичная структура белка определяется сверткой, в которой число пар гидрофобных аминокислот, соответствующих смежным вершинам графа и не являющихся соседними членами первичной структуры, максимально. Такие пары называются контактными парами (или просто контактами). В дальнейшем, без ограничения общности будем считать, что белок задан в виде бинарной последовательности, где единицам соответствуют гидрофобные аминокислоты, а нулям – гидрофильные.

Пусть дана строка $p = p_1 p_2 \dots p_m$, $p_i \in \{0, 1\}$, $i = \overline{1, m}$ и граф G с множеством вершин $V(G)$ и множеством ребер $E(G)$. Инъекция $\varphi: \{1, 2, \dots, m\} \rightarrow V(G)$, такая, что $\{\varphi(i), \varphi(i+1)\} \in E(G)$, $i = \overline{1, m-1}$, называется укладкой строки p в граф G . Укладка называется конформацией, если никакие два ребра в укладке не имеют геометрического пересечения. Таким образом, задача о свертке белка в рамках НР-модели сводится к нахождению конформации с наибольшим числом контактов среди всех возможных конформаций. Известно [6], что эта задача является NP-трудной для графов, ассоциированных с двумерной квадратной и трехмерной кубической решетками.

В работе исследована задача о свертке белка в случае триангулированной двумерной и трехмерной решеток, а также квадратной двумерной решетки с диагоналями в рамках стандартного НР-подхода. Для этой задачи разработаны приближенные алгоритмы линейной (от длины строки) сложности и обладающие гарантированными оценками точности:

- алгоритм остовной укладки для двумерной триангулированной решетки с оценкой 4;
- алгоритм укладки стрелой для двумерной триангулированной решетки с оценкой 2;
- алгоритм укладки звездой для трехмерной триангулированной решетки с оценкой 1.875;
- алгоритм остовной укладки для двумерной квадратной решетки с диагоналями с оценкой 3.

При исследовании задачи в случае триангулированной решетки использованы локальные правила укладки строк из [4]. На их основе разработаны и программно реализованы соответствующие приближенные алгоритмы. Более того, уточнены и доказаны утверждения, с помощью которых получаются оценки точности алгоритмов. Все доказательства ввиду ограниченности объема статьи не приводятся.

Графом двумерной триангулированной решетки назовем граф $G = (V(G), E(G))$, для которого выполняются следующие условия:

$$V(G) = \left\{ \left(x + \frac{y}{2}, \frac{y\sqrt{3}}{2} \right) \mid x, y \in \mathbb{Z} \right\},$$

$$E(G) = \{ \{u, v\} \mid u, v \in V(G), \|u - v\|_2 = 1 \}.$$

Такая решетка состоит из равносторонних треугольников с длиной стороны, равной 1. Под выражением $\|u - v\|_2$ понимается норма в евклидовом двумерном пространстве разности двух векторов.

Остовный алгоритм для двумерной триангулированной решетки является 4-приближенным алгоритмом и представляет собой очень простую стратегию укладки, основанную на создании двух параллельных остовов. Остовами называются две выбранные соседние параллельные прямые, с одной стороны от которых вершины укладываются “елочкой”. В процессе укладки строка p разбивается на подстроки вида $0^i 1^j$, где $i, j > 0$. Основные правила укладки подстроки $0^i 1^j$ приведены на рис. 1. Эти правила формируют конформацию, удовлетворяющую двум условиям: все единицы лежат на каком-нибудь из остовов; два нуля на одном остоле никогда не могут быть соседними.

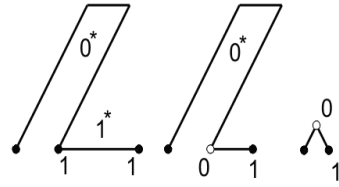
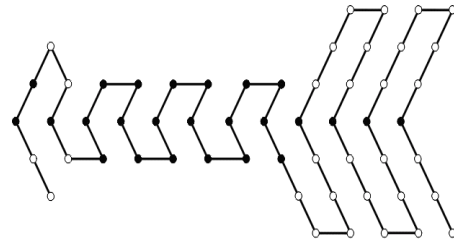


Рис. 1. Основные правила укладки подстроки вида $0^i 1^j$ (1^* обозначает $(j - 2)$ единицы)

Алгоритм укладки стрелой является 2-приближенным алгоритмом и основан на выборе трех параллельных соседних прямых, которые называются уровнями. На трех уровнях лежат все единицы. На среднем уровне расположены только единицы. Такая укладка строки достигается за счет того, что мы выделяем в строке максимальную подстроку, состоящую из одних единиц (так называемый стрип) и укладываем ее “змейкой” проходя через три уровня последовательно. Точная укладка стрипа зависит от того, какой остаток имеет его длина при делении на 3 и направления движения по уровням. На рис. 2 приведен пример укладки стрелой.



Графом трехмерной триангулированной решетки назовем граф $G = (V(G), E(G))$, для которого выполняются следующие условия:

$$V(G) =$$

$$\left\{ \left(x + \frac{y}{2} + \frac{z}{2}, \frac{y\sqrt{3}}{2} + \frac{z}{2\sqrt{3}}, z\sqrt{\frac{2}{3}} \right) \mid x, y, z \in \mathbb{Z} \right\},$$

Рис. 2. Пример укладки стрелой для последовательности

$p = 001100101111111111111111111110$
 $00001000000110000010000001000$

$$E(G) = \{ \{u, v\} \mid u, v \in V(G), \|u - v\|_3 = 1 \}.$$

Такой граф состоит из равносторонних тетраэдров с длиной стороны, равной 1 и получается путем параллельного переноса триангулированных двумерных решеток вдоль оси Oz на величину $\sqrt{2/3}$, относительно предыдущего слоя и сдвига в плоскости Oxy .

Суть алгоритма укладки звездой состоит в следующем. Помещаем все единицы в одну плотноупакованную гидрофобную (стабильную) структуру. Основа такой структуры составлена не более чем из шести единиц на каждом уровне. Такая структура составлена из уровней, каждый из которых представляет фрагмент двумерной триангулированной решетки. Показано, что алгоритм укладки звездой имеет оценку приближения, равную 1.875.

Графом квадратной двумерной решетки с диагоналями назовем граф $G = (V(G), E(G))$, для которого выполняются следующие условия:

$$V(G) = Z^2 = Z \times Z,$$

$$E(G) = \{ \{u, v\} \mid u, v \in V(G), \|u - v\|_2 \leq \sqrt{2} \}.$$

Иллюстрация работы алгоритма остовной укладки для двумерной квадратной решетки с диагоналями приведена на рис. 3. Доказано, что остовный алгоритм для данной решетки является 3-приближенным.

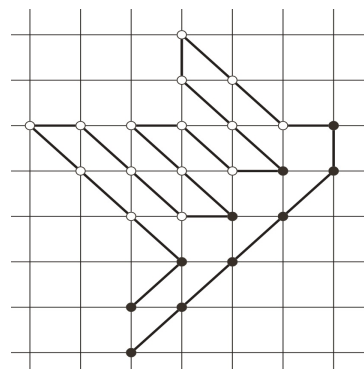


Рис. 3. Пример остовной укладки для двумерной квадратной решетки с диагоналями

Литература

1. *Branden C.-I., Tooze J.* Introduction to protein structure / Garland Publishing. 1999.
2. *Baxevanis A., Francis Ouellette B.F.* Bioinformatics: a practical guide to the analysis of genes and proteins / John Wiley & Sons, Inc. 1998.
3. *Rashidi H.H., Buehler L.K.* Bioinformatics basics applications in biological science and medicine / CRC Press. 1999.
4. *Agarwala R., Batzoglou S., Dan V., Decatur S.E., Farach M., Hannenhalli S., Skiena S.* Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the HP model / J. Comput. Biology. 1997. Vol. 4. P. 275–296.
5. *Dill K.A.* Theory for the folding and stability of globular proteins / Biochemistry. 1985. Vol. 24. P. 1501.
6. *Crescenzi P., Goldman D., Papadimitriou C., Piccolboni A., Yannakakis M.* On the complexity of protein folding / J. Comput. Biology. 1998. Vol. 5. P. 423–466.

РАСПАРАЛЛЕЛИВАНИЕ АЛГОРИТМА ЧИСЛЕННОГО РЕШЕНИЯ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ ПЕРВОГО ПОРЯДКА

А. Ю. Лобанов

В связи с тем, что физические возможности совершенствования вычислительных устройств ограничены, в последнее время многие надежды на увеличение производительности связаны со специализированными